

April 1994

## Rawls's New Theory of Justice

Rex Martin

Follow this and additional works at: <https://scholarship.kentlaw.iit.edu/cklawreview>



Part of the [Law Commons](#)

---

### Recommended Citation

Rex Martin, *Rawls's New Theory of Justice*, 69 Chi.-Kent L. Rev. 737 (1994).

Available at: <https://scholarship.kentlaw.iit.edu/cklawreview/vol69/iss3/8>

This Article is brought to you for free and open access by Scholarly Commons @ IIT Chicago-Kent College of Law. It has been accepted for inclusion in Chicago-Kent Law Review by an authorized editor of Scholarly Commons @ IIT Chicago-Kent College of Law. For more information, please contact [dginsberg@kentlaw.iit.edu](mailto:dginsberg@kentlaw.iit.edu).

# RAWLS'S NEW THEORY OF JUSTICE

REX MARTIN\*

## I. BACKGROUND

In 1971 John Rawls published his monumental treatise, *A Theory of Justice*. The book has been widely discussed, both favorably and unfavorably. As I reflected on the book, I concluded that it had at least two main problems. Let me sketch the most significant of these problems and, then, merely mention the other.

The complicated procedure whereby Rawls attempted to justify his two principles of justice—a procedure centering on the notion of an “original position” for deciding about candidate principles of justice—has, I think, important defects. For one thing, it merely assumed some of the very things (the values of liberty and equality, for example) that a sound *critical* justification would need to vindicate, not assume. And, for another, in making these very assumptions, Rawls's theory seemed to rule out from serious consideration certain rival candidates to his own two principles (especially those candidates, such as Platonic perfectionism or Nietzschean elitism, that did not take the equality or liberty of individuals as fundamental).

Rawls had stacked the deck in the original position against these perfectionist theories. For, arguably, the only principles of justice that could survive scrutiny by the “parties”—by the body of fellow citizens (or their representatives), each one of them having equal status and an equal voice and full veto power—are principles that treat people as substantive equals.

The point, then, is simply that Rawls's theory, given the defects I have outlined in the assessment procedure, is not, *on its own terms*, an acceptable or accredited theory of critical moral justification. For that procedure does not satisfy its own goal: of wielding a set of *objectively* based considerations for fairly assessing rival principles of justice.

\* Professor of Philosophy, University of Kansas. For contributing to my thinking on the themes of the present Article, I want to acknowledge the help provided by Philip Kissam and Russ Shafer-Landau and the other members of my reading group on *Political Liberalism* (during 1993-94), by the written comments of Ted Vaggalis and of David Reidy on an earlier draft of this paper, and by the comments of various audiences in Australia, New Zealand, and Wales on even earlier incarnations of it.

My second criticism, a lesser one, is that Rawls did not develop a compelling account of the content of his first principle—that is, he did not lay out a convincing account of why some liberties should be taken as basic. Nor did he develop any sort of plausible argument to the effect that those liberties should have priority over other elements in the theory of justice. His further claim that they have priority over such values as the common good (that is, general welfare) or efficiency or the perfectionist values mentioned earlier is left in an almost equally unsatisfactory position.<sup>1</sup>

To this list of failings, Rawls himself has added one of his own. He claims that his basic account of stability in *A Theory of Justice* is all wrong. Let me put the point of demur crudely here: Rawls's assumption in 1971 was that since the two principles came out on top in the contest with utilitarianism and with perfectionist values, the two principles would become the moral theory or part of the moral theory of any well-ordered society in which these principles were the public principles of justice. Such a society would be stable, in short, because everybody in it would adopt the two principles as among their moral principles.

But this uniform acceptance, Rawls says, is inconsistent with the idea that a pluralism of reasonable comprehensive moral doctrines is here to stay. The pluralism of moral principles is a permanent and ineradicable fact of modern political society.

It seems, then, that Rawls's operating assumption (in 1971) goes against the subtext, against an underlying but unannounced theme of *A Theory of Justice* in favor of this ineradicable pluralism. Or, if not that, the assumption is inconsistent with what Rawls *now* regards as the most realistic of foreseeable conditions under which those principles could be realized.<sup>2</sup> Thus Rawls's assumption (in 1971) was either self-contradictory (given the text and subtext of *A Theory of Justice*)

1. Both these failures are acknowledged in John Rawls, *The Basic Liberties and Their Priority*, in 3 THE TANNER LECTURES ON HUMAN VALUES 3-87 (Sterling M. McMurrin ed., 1982) [hereinafter Rawls, *Basic Liberties*]. For his acknowledgment of the second failure (the one concerned with priority), see JOHN RAWLS, *POLITICAL LIBERALISM* 369-71 (1993).

2. The first of these options was suggested to me by Philip Kissam. For one can plausibly read chapter four of *A Theory of Justice* as premised on the idea that the *need* for liberty of conscience is permanent, since the condition of an irreducible diversity of fundamental religious and moral and philosophical beliefs is itself permanent. JOHN RAWLS, *A THEORY OF JUSTICE* (1971) [hereinafter RAWLS, *THEORY*]. The second of these options (the one concerned with giving expression to the most realistic of foreseeable conditions) was suggested to me by Russ Shafer-Landau. For Rawls's own account of the matter, see RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at xv-xvii.

or it was inconsistent with the best available insights as to how stability could be realized.

Rawls now seems most impressed with this third problem, the problem of assuring stability. In any event, he clearly became dissatisfied with the shape his theory had taken in 1971 and had retained for about a decade afterwards. He began to rethink that theory. For Rawls seemed, beginning with his Dewey Lectures in 1980, to be reconfiguring his entire justificatory account. A number of important changes have occurred as he has moved further from positions he occupied in *A Theory of Justice*.

## II. RAWLS'S NEW THEORY

Rawls now claims that his theory is specifically a *political* theory of justice, which is itself not a comprehensive critical moral theory, *nor a part of one*. Or at least it is a part of no one such theory in particular. Rather, in this newer account, such reasonable comprehensive and general moral theories (as Kant's ethical theory or as Mill's utilitarianism) merely "overlap" on the independently justified "political conception of justice."<sup>3</sup> And the values that Rawls uses to construct this independent justification of the principles of justice ingredient in that conception are themselves said to be latent in the public political culture of a contemporary democratic society.

When I speak of Rawls's "new theory of justice" it is this increasingly *political* theory of justice that I have in mind. Intimations of a significant change here were in view as early as his Dewey Lectures. But the first essay to make the scale of intended revision quite clear

3. The phrase "comprehensive and general" is from Rawls. Critical theories are general, Rawls says, when they apply to "a wide range of subjects of appraisal (in the limit of all subjects universally)," and comprehensive when they include "conceptions of what is of value in human life, ideals of personal virtue and character, and the like, that are to inform much of our conduct (in the limit our life as a whole)." John Rawls, *The Idea of an Overlapping Consensus*, 7 OXFORD J. LEGAL STUD. 1, 3 n.4 (1987) [hereinafter Rawls, *Overlapping Consensus*]. He adds that, "[m]any religious and philosophical doctrines tend to be general and fully comprehensive." *Id.*

For further discussion and examples, see RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 13, 152 n.17; John Rawls, *The Domain of the Political and Overlapping Consensus*, 64 N.Y.U. L. REV. 233, 240 (1989) [hereinafter Rawls, *Domain*]; John Rawls, *The Priority of Right and Ideas of the Good*, 17 PHIL. & PUB. AFF. 251, 252-53, & n.2 (1988) [hereinafter Rawls, *Priority of Right*]; Rawls, *Overlapping Consensus*, *supra*, at 6, 14 n.23. For the idea that such theories are or might be reasonable, see RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 58-66, and for the claim that the political conception is not itself a comprehensive doctrine, or a part of one, see *Id.* at 154-58, 175.

I should add that, in the end, Rawls regards most actual critical moral theories, though comprehensive and general in intent, as only *partially* comprehensive and general. See *id.* at 168, 208; Rawls, *Priority of Right*, *supra*, at 274-75; Rawls, *Overlapping Consensus*, *supra*, at 19. Thus, we should properly speak of such theories as *relatively* comprehensive and general.

was his article *Justice as Fairness, Political Not Metaphysical*.<sup>4</sup> And his essays since then have formulated this new political theory of justice in ever finer detail.<sup>5</sup>

These essays have been brought together in his new book, *Political Liberalism*. As might be expected, Rawls's new theory is given its most complete elaboration in this book. The first part of the book is an extensive reworking of Rawls's three Dewey Lectures.<sup>6</sup> In homage to the origins of this part of the book, and in trying to adopt a more conversational tone than he had taken in his 1971 treatise, Rawls refers to the main divisions of his new book, not as chapters, but as lectures. A closer look reveals that the first of the original Dewey Lectures has more or less been supplanted by a revised version of *Justice as Fairness, Political Not Metaphysical*. Lectures II and III in the book are revised versions of the next two of the original Dewey Lectures; these revisions were introduced largely to take account of the *political* character of his new theory and of the constraints this particular conception imposes.

The second part of the book consists of three more lectures which weave together various parts of his recent articles<sup>7</sup> and some new, previously unpublished material on "public reason." This new material is by and large found in Lecture VI. The final part of the book consists of two chapters (again called lectures); these are simply republications, unchanged, of previous articles. Thus, Lecture VII in the book is Rawls's 1978 article entitled *The Basic Structure as Subject*,<sup>8</sup> and Lecture VIII is his Tanner Lecture, *The Basic Liberties and Their Priority*.<sup>9</sup> To this rather unwieldy amalgam of lectures and articles and other material Rawls has added a new and very interesting introductory essay.<sup>10</sup>

How might we best picture this new theory? Perhaps, its most significant feature is that Rawls takes the public political culture of a contemporary democratic society to be the deep background of the

4. John Rawls, *Justice as Fairness: Political Not Metaphysical*, 14 PHIL. & PUB. AFF. 223 (1985).

5. See, e.g., Rawls, *Domain*, *supra* note 3; Rawls, *Priority of Right*, *supra* note 3; Rawls, *Overlapping Consensus*, *supra* note 3.

6. John Rawls, *Kantian Constructivism in Moral Theory: The Dewey Lectures 1980*, 77 J. PHIL. 515 (1980).

7. See, e.g., Rawls, *Domain*, *supra* note 3; Rawls, *Priority of Right*, *supra* note 3; Rawls, *Overlapping Consensus*, *supra* note 3.

8. John Rawls, *The Basic Structure as Subject*, in VALUES AND MORALS 47 (Alvin I. Goldman & J. Kim eds., 1978).

9. Rawls, *Basic Liberties*, *supra* note 1.

10. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at xiii-xxxiv.

entire theory. For the leading ideas out of which the political conception of justice is to be constructed and by reference to which it is to be justified are said by Rawls to be implicit in that culture.

More specifically, political justification, in Rawls's post-1980s writings, sets out from four "model conceptions" or "fundamental ideas": the idea of the person or citizen, the idea of social cooperation for reciprocal benefit, the idea of the well-ordered society and its basic structure of institutions, and the idea of a linking or mediating conception which lays out the standards for discussion and for decision making that fellow citizens could be expected to follow in reaching a rational and reasonable decision respecting the governing principles of political justice (that is, the principles of justice for the basic structure of their well-ordered society). Not surprisingly, this mediating conception captures many of the features familiar from Rawls's earlier discussion of the original position (including, of course, many of the restrictions imposed by what he calls the "veil of ignorance"). Technically, the mediating conception is not implicit *per se* in democratic culture; rather, its function is to help unify the other fundamental ideas into a single coherent whole from which one could then reason, with some effect, to certain principles and institutional arrangements.<sup>11</sup>

Establishing some such set of principles—principles for a fair distribution of certain primary goods (liberties, opportunities, social and economic positions, income and wealth)—is the main object of Rawls's new *political* conception of justice. In this account, the princi-

11. See RAWLS, *THEORY*, *supra* note 2, at 118-92 (especially pages 126-27, 137, 146-47), for his most complete account of the so-called original position. Perhaps, the leading idea in this account is the notion that the "parties" decide about candidate principles of justice behind a "veil of ignorance." See *id.* at 136-42. The "veil" here is metaphorical. It stands for the idea that certain identifiable, particular facts about the parties (for example, their strength, intelligence, religion, determinate conception of the good) are bracketed off from the discussion, as being prejudicial or irrelevant matters. Here, then, the parties are not allowed to rely on these facts, or to appeal to them in effect, in their arguments or in the decision they make about which candidate principle is to be preferred. In this regard, the veil functions like the instructions a judge makes to a jury to overlook or not consider certain points in their deliberations.

In his later writings, as I said, Rawls treats the original position as one of several "model-conceptions." The original position model is explicitly described as the "mediating" conception between other, more fundamental conceptions, see RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 304-05; Rawls, *Basic Liberties*, *supra* note 1, at 18-19; Rawls, *supra* note 6, at 520, 522, 533, 566, or as a "device of representation," a device for representing certain considerations as regards moral persons or their cooperation for mutual benefit. See Rawls, *supra* note 4, at 234-39. And in his most recent articles, the original position idea is mentioned only perfunctorily and relegated to the footnotes. See Rawls, *Domain*, *supra* note 3, at 247 n.32, 251 n.46; Rawls, *Priority of Right*, *supra* note 3, at 274 n.32; Rawls, *Overlapping Consensus*, *supra* note 3, at 7 n.13.

Nonetheless, that model continues to play a modest but necessary role in Rawls's theory, as I have indicated. See, e.g., RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 22-28; see also *id.* at xxix, 14 n.16, 72, 75, 90, 103, 106-7, 207, 227.

ples that emerge as preferred (from among a small set of historically available candidate principles) are, presumably, the principles that are best supported from within the nexus formed by the model conceptions. The preferred principles are the ones that, upon reflection and given the balance of reasons, are the most appropriate ones with respect to the starting point itself. That is, they are the principles most appropriate to the fundamental ideas themselves under the presumption that there is and is going to be an irreducible pluralism of reasonable comprehensive moral and religious doctrines.

Thus, fellow citizens, acting in accordance with the norms of discourse and of decision-making specified in the mediating model, are able rationally to affirm a shared set of principles of justice—principles that they would prefer to see used to organize a unified and stable scheme of basic political and economic institutions under which they can all live together. The principles so selected are not principles for the whole of life; they are not comprehensive and general moral principles; rather, they are merely the appropriate principles for socially distributing primary goods and for governing the life of fellow citizens *in the public domain*, coercively if need be.

Rawls thinks that those best-supported principles will be his own two principles of justice, understood now as *political* principles. But the “political conception of justice,” as Rawls calls it, is not limited to these principles alone. It also includes certain of the institutional arrangements that are required to put the principles into effect in a given society. These institutions—political, economic, social—are the sort of thing Rawls had in mind when he referred to the basic structure of a society. But he tends to emphasize only the basic political institutions (the “constitutional essentials,” as he calls them).<sup>12</sup> This is another sense, then, in which the political conception of justice is *political*. For the conception not only takes its rise and finds initial justification for its embedded principles from ideas latent in the public political culture of a contemporary democratic society, it also includes some of the political institutions to be found in the basic structure of such a society.

12. For the phrase “constitutional essentials,” see RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at xxi, and for more extended discussion of these essentials, see *id.* at 227-40. For Rawls’s discussion of the political conception of justice, see *id.* at 10, 11-15, 64, 71, 103. Rawls explicitly includes the constitutional essentials within the political conception in *id.* at 128 and 156. For his frequent bracketing of such essentials together with “basic justice,” see *id.* at 137, 140, 156, 190, 214, 215, 224, 241, 244 n.33.

Rawls claims that this political conception will be endorsed and supported as the focus of an "overlapping consensus" among the various comprehensive religious and moral doctrines that exist in the Western world today. Why is the idea of overlapping consensus brought into play by Rawls?

Three reasons can be offered. First, Rawls thinks that, unless his political theory can be critically morally justified, it will strike people as a mere *modus vivendi*. I would point out that it is already more than that: it is a carefully crafted vindication of political institutions as *just* by reference to standards actively embedded within a living political culture, to standards and practices that would be reflectively acknowledged as sound by a wide spectrum of members of that culture. And reflectively acknowledged by them as being in the interest of each and all, as being for their common good. (It is even possible that such standards and practices would be reflectively acknowledged as sound, or at least reasonable, by persons from other cultures or different times as well, but this is not Rawls's main concern.)

Rawls thinks this vindication, solid as it is on its own terms, lacks deep moral credentials (of the sort afforded by a comprehensive critical moral theory). Thus, one point (in Rawls's view) of critical moral justification would be to *moralize* the *modus vivendi*, to give the political justification the moral grounding it otherwise lacks.<sup>13</sup>

Rawls characteristically describes the critical moral justification in question as being one of overlapping consensus. This reflects his view that no single comprehensive critical moral theory could be expected to support the politically justified system in every detail. More important, it reflects his view that that system could be partially supported by several different critical moral theories, each of them controversial, no one of them accepted by everybody (or almost everybody), and all of them subject to endless and apparently unresolvable disputation.

The point of multiple support here is not to moralize the system even further (by taking in every detail, covering every base, from one angle or another). Rather, the point of multiple support by a variety of accredited (or at least widely acknowledged) critical moral or religious theories is to lend greater stability, over time and for the long run, to the political system itself. Thus, Rawls here takes explicit ac-

13. For this point (about moralizing the grounds of the political conception), see RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 126, 144-50, 168, 208; Rawls, *Priority of Right*, *supra* note 3, at 274; Rawls, *Overlapping Consensus*, *supra* note 3, at 2, 11; Rawls, *Justice as Fairness*, *supra* note 4, at 247.



count of the basic fact of the permanent and ineradicable pluralism of reasonable moral and religious conceptions, and he fashions a theory of realizing stability in the system that is consistent with this fact (as he had self-professedly failed to do in *A Theory of Justice*).<sup>14</sup> The realization of stability is the second reason Rawls had for introducing overlapping consensus into his theory of the political conception of justice.

Finally, Rawls wants to insure, with overlapping consensus, that the system will have and continue to enjoy widespread support from its citizens, regardless of their ethical commitments and regardless of whether they have internalized the political justification in its own terms. For here *other* terms of justification and of obligation to laws and political arrangements—those provided by the various competing moral and religious theories—are readily available to *these* citizens.<sup>15</sup>

Rawls thinks, if we descend now to specifics, that several such supports can be afforded his politically justified system. Among them (as I already mentioned) are two well-known and presumably accredited theories of critical moral justification: Kantian ethical theory and Mill's utilitarianism.<sup>16</sup>

Rawls's new theory has a vast number of interesting turns and wrinkles. I will be concerned with just one of them in the next section:

14. See *supra* part I.

15. Two considerations are raised in this and the previous paragraph of the text: (a) the role of overlapping consensus in affording stability, RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 65-66, 134, 140-44, 148; Rawls, *Priority of Right*, *supra* note 3, at 275; Rawls, *Overlapping Consensus*, *supra* note 3, at 10-11, 22-24; Rawls, *supra* note 4, at 250-51, and (b) the attendant provision of support by the citizens on grounds other than the political conception itself. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 147-48, 150, 187; Rawls, *Priority of Right*, *supra* note 3, at 269.

16. See RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 145; Rawls, *Overlapping Consensus*, *supra* note 3, at 6 & n.11, 9; Rawls, *supra* note 4, at 245 & n.28, 246, 250. Rawls often refers to these theories as the "liberalisms of Kant and Mill." See, e.g., RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 135 n.3, 145, 159, 199, 200, 211 n.42; Rawls, *supra* note 4, at 6.

It is clear that when Rawls cited Mill's theory, as possibly part of the overlapping consensus, he particularly had in mind chapter three of Mill's *On Liberty* as the locus classicus of Mill's "liberalism" (the term Rawls normally used to refer to Mill's theory and to group it with Kant's). See JOHN STUART MILL, *ON LIBERTY* 53-69 (David Spitz ed., W.W. Norton & Co. 1975) (1859). He does say at one point that Mill's liberalism is not a form of utilitarianism. John Rawls, *Social Unity and Primary Goods*, in *UTILITARIANISM AND BEYOND* 159, 160 (A. Sen & B. Williams eds., 1982). But the more important point here is that in his most recent articles Rawls does not exclude the possibility that Mill's utilitarianism, suitably interpreted, could (unlike Benthamite or classical utilitarianism) be part of the overlapping consensus that supports the Rawlsian political conception of justice. See Rawls, *Domain*, *supra* note 3, at 249 & n.39; Rawls, *Overlapping Consensus*, *supra* note 3, at 12 & nn.18, 20.

In *Political Liberalism*, however (and in keeping with the more inclusive tone struck in that work), Rawls asserts that even classical utilitarianism (the utilitarianism of Bentham and Sidgwick) could be seen to be part of that consensus. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 169-71.

his current formulation of the two principles of justice, in particular, the first principle. Then, in the section after that, I want to return to the larger issue of justification by sketching out problems I see for Rawls's account in each of its two main stages.

### III. REFORMULATION OF THE TWO PRINCIPLES OF JUSTICE

Rawls's current version of the two principles of justice reads as follows:

- [1.] Each person has an equal claim to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with the same scheme for all; and in this scheme the equal political liberties, and only those liberties, are to be guaranteed their fair value.
- [2.] Social and economic inequalities are to satisfy two conditions: first, they must be attached to positions and offices open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least advantaged members of society.<sup>17</sup>

This formulation is different in important respects from the one Rawls had used in *A Theory of Justice*.<sup>18</sup> The main difference comes in the first principle. Thus, in 1971 Rawls spoke of "the most extensive total system of equal basic liberties";<sup>19</sup> whereas in 1993 that phrase is replaced by "a fully adequate scheme of equal basic rights and liberties."<sup>20</sup> Rawls introduced this new language in his Tanner Lectures and, of course, restructured his supporting arguments there, in response to criticisms by Hart.<sup>21</sup> That new language was repeated thereafter.<sup>22</sup>

When Rawls first states his two principles in *Political Liberalism*, he notes that his formulation there "follows" the one given in the Tanner Lectures. But it is not identical; several changes have occurred, in particular, in the first principle.

In Tanner (as in *A Theory of Justice*) Rawls roots the entire first principle in the notion of a fundamental right; thus, the formulation begins by saying "each person has an equal right," and the remainder of the first principle is given, then, as the *content* of that right. But in *Political Liberalism* he says, not that each has an "equal right," but

17. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 5-6.

18. As one can tell by a glance at the canonical version he provides in RAWLS, *THEORY*, *supra* note 2, at 302.

19. *Id.*

20. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 5.

21. See *id.* at 291; Rawls, *The Basic Liberties*, *supra* note 1, at 5.

22. See Rawls, *Domain*, *supra* note 3, at 251 n.43; Rawls, *supra* note 4, at 227.

that each has an "equal claim." And in the next line of *Political Liberalism* he speaks of "equal basic rights and liberties" (though in Tanner and in *A Theory of Justice* he had made do merely with "equal basic liberties"). Rawls neither notes these changes nor adumbrates reasons (or even hints) as to why he had made them. One does not know then what to make of Rawls's shuffling about of the terms 'claims' and 'rights.'<sup>23</sup>

One other change occurs in the formulation of the first principle between Tanner and *Political Liberalism*. I mean the addition of the explicit directive (and attendant disclaimer) in *Political Liberalism* that "in this scheme the equal political liberties, and only those liberties, are to be guaranteed their fair value." This language does not occur in the Tanner formulation of the first principle. But the argument for singling out "the equal political liberties" and guaranteeing "their fair value" is made at some length in Tanner. So the difference here (between the formulations in Tanner and *Political Liberalism*) is more apparent than real, though it marks a real difference between those two works and *A Theory of Justice*.<sup>24</sup>

It is interesting to note, almost as an aside, that the formulation of the *second* principle is virtually unchanged between *A Theory of Justice* and *Political Liberalism*. This would suggest that Rawls has

23. Nor can one be at all certain of the importance of attending to such changes. However, I can suggest some reasons why these changes might be thought desirable. Thus, in my book *Rawls and Rights*, I provide some arguments as to why Rawls might not want to state his first principle of justice in such a way as to suggest that it is itself a noninstitutional natural right. REX MARTIN, *RAWLS AND RIGHTS* 21-41 (1985).

And I argue in that same book that several of the things Rawls discusses under the heading of liberties are not things an agent does or can do but, rather, injuries to be avoided (at the hands of others). *Id.* at 46-52. Thus, these noninjuries, though rights in his account, are not liberties. And Rawls's new language (of "equal basic rights and liberties") could be designed to accommodate such noninjuries within his first principle of justice. *RAWLS, POLITICAL LIBERALISM, supra* note 1, at 5.

24. Rawls, in the context of arguing for the equal political liberties, has claimed that individual citizens should be guaranteed "fair value" there—that is, a rough equality of access to political participation. Rawls, *Basic Liberties, supra* note 1, at 39-46, 75-78; *see also* *RAWLS, POLITICAL LIBERALISM, supra* note 1, at 324-31, 359-62. "[T]his guarantee means," he says, [t]hat the worth of the political liberties to all citizens, whatever their social or economic position, must be approximately equal, or at least sufficiently equal, in the sense that everyone has a fair opportunity to hold public office or to influence the outcome of political decisions.

*Rawls, Basic Liberties, supra* note 1, at 42; *see also* *RAWLS, POLITICAL LIBERALISM, supra* note 1, at 327-28. And this equalizing of the opportunity to count politically is to be achieved through such measures as (a) limits on campaign spending, or (b) government funding of campaigns, or (c) public subsidy (or a mandate for ready and inexpensive availability of air time) to candidates and political parties and other relevant points of view. His point is there must be allowance for the real input of individual persons in the political process, at "fair value," in a democratic scheme of citizenship. *RAWLS, POLITICAL LIBERALISM, supra* note 1, at 235 n.22. For further discussion, *see* *MARTIN, supra* note 23, at 45-61, 139-40, 160, 217 n.24.

been far more concerned, in the thinking that has gone into his new theory of justice, with reworking the first principle of justice than with reworking the second.

Indeed, the only change worth noting in his latest formulation of the second principle is merely cosmetic. In *A Theory of Justice* Rawls had mentioned "the greatest benefit" before he mentioned the idea of "fair equality of opportunity." In both Tanner and *Political Liberalism* he reverses this order of presentation, putting (a) "fair equality of opportunity" ahead of (b) "greatest benefit" (which he refers to elsewhere as the difference principle or, sometimes, the maximin principle).

This is not an important change. But it, nonetheless, has one virtue. The way Rawls now states his two principles of justice (in Tanner and *Political Liberalism*) reflects and conforms to the relevant internal priorities among the constituent elements:  $1 > 2a$ ,  $1$  or  $2a > 2b$ .<sup>25</sup>

However, it continues to be the case that Rawls does not make an adequate case for the priority of the first principle of justice over the second—of certain basic rights over elements in the second principle, of those fundamental rights over policies designed to achieve fair equality of opportunity or to achieve maximization of the level of goods and services available to the least well-off income group (say, the bottom 20%). Indeed, for that matter, we cannot say even that Rawls makes a satisfactory case for putting basic constitutional rights (here, specific basic rights to liberty and noninjury) *over* policies advancing corporate goods or other aggregate considerations. This, then, is a continuing crucial failure of Rawls's theory insofar as it attempts to provide a justification of the standing priority of basic constitutional rights over such considerations.

Other problems press forward as well, in particular questions about the justification of the two principles. I want now to turn to some of those questions.

25. For these priorities, see RAWLS, *THEORY*, *supra* note 2, at 61, 87-89, 302-03. In *Political Liberalism*, Rawls tends to state the *internal* priorities within his two principles more simply. Here he alleges only the priority of the first principle over the second. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 6. Or sometimes only the priority of the first principle plus certain liberties that come under fair equality of opportunity (such liberties as "freedom of movement and free choice of occupation") over the remainder of the elements in the second principle. *Id.* at 228-29.

#### IV. PROBLEMS OF JUSTIFICATION

It is convenient to view Rawls's account of justification in *Political Liberalism* as having two main stages. In the first stage Rawls provides a justification of the political conception of justice by drawing on certain fundamental ideas which he finds implicit in the public political culture of a contemporary democratic society. The justification here is said to be "freestanding" in the sense that it draws only on these implicit ideas, presumably shared already to a large degree by fellow citizens; it does not draw on the ideas or values of any comprehensive moral or religious doctrine *per se*.

In the second stage, then, Rawls attempts to show that this political conception will also be endorsed and supported as the focus of an overlapping consensus among comprehensive religious and moral doctrines. On this view, the political conception is a "module, an essential constituent part, that fits into and can be supported by various reasonable comprehensive doctrines that endure in the society regulated by it."<sup>26</sup> But it is not *presented* by reference to such support initially; rather it is established independently of direct consideration of any and all such doctrines. Even at the second stage, it is not regarded as a part of any one of them in particular (to the exclusion of all or most others).

In my judgment Rawls's idea of two-stage justification, as just summarized, is one of the most distinctive and promising differences between *A Theory of Justice* and his new theory of justice (as it is definitively stated in *Political Liberalism*). In my discussion of Rawlsian justification, I will follow this leading idea, of two stages of justification. I will take the stages up in turn, beginning with the freestanding or political one.

##### A. *Political Justification in a Democratic Setting*

Among the features of Rawls's political conception of justice, as we have just noted, is the crucial claim that all the background elements (in particular, the fundamental ideas) that ground the justification of the substantive principles of justice are themselves latent or implicit in the practices and attitudes of a contemporary democratic society.<sup>27</sup> In *Political Liberalism*, the original position model (or me-

26. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 12. See also *id.* at 145.

27. See, e.g., RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 13, 175, 223; Rawls, *Domain*, *supra* note 3, at 240; Rawls, *Priority of Right*, *supra* note 3, at 252; Rawls, *Overlapping Consensus*, *supra* note 3, at 6, 8; Rawls, *supra* note 4, at 231 n.14.

diating conception, as Rawls now tends to call it) is a heuristic device for ordering these pre-existing fundamental ideas (ideas of society as a cooperative scheme for reciprocal benefit, of the citizens as free and equal, and so on) into a single coherent whole. To do that and to regiment these ideas so as to be able to exhibit, on the balance of reasons, which among the historically available candidate principles of justice is the one most suitable to those ideas. The main project here (as Rawls now sees it) is to settle on that principle or set of principles for distributing primary goods which is optimally appropriate, given the fundamental democratic ideas from which we start.

What Rawls actually does, however, is akin to but not identical with the program just outlined in the previous paragraph. He argues, more or less straightforwardly, that the two principles, when reworked (as described in section III above), are peculiarly well designed to specify a democratically acceptable distribution of primary goods in the context of existing democratic political arrangements. His project so understood is an interpretive one. Its main object is to *construct* some such optimal set of principles.<sup>28</sup> But we are no longer required to carry out this project comparatively (by showing, for example, that the two principles are to be preferred from among the set of *historically* available ones). Thus, relatively little time is spent (in the new theory) on disputing the merits of the principles that compete with the two principles. Indeed, doing so might prove counterproductive, once we move to the *next* stage—that of overlapping consensus.<sup>29</sup>

But I do not think Rawls has adequately attended to the change he has wrought in his own theory at this point. In *A Theory of Justice*, he had generated the various political practices, which collectively add up, roughly, to those included under the heading of democratic institutions, out of features of the original position.<sup>30</sup> But in the new political conception of justice, the original position model can no longer be used to generate the so-called democratic institutions and their attendant values or to justify them. For the original position model nec-

28. Rawls makes this clear in *POLITICAL LIBERALISM*, *supra* note 1, at 89-129.

29. I owe the germ of this idea to Ted Zenzinger. There are, I believe, two reasons for this shift in Rawls's approach. First, it is fair to say (as students in one of my seminars observed) that Rawls has not developed a political conception for utilitarianism or for perfectionism to set alongside his own political conception of justice. Second, in his new mood of conciliation (as found in *Political Liberalism*), Rawls is reluctant to argue against *any* reasonable conception of the good and, hence, reluctant to argue for his two principles by arguing *against* utilitarian or perfectionist alternatives.

30. See RAWLS, *THEORY*, *supra* note 2, at 221-34, 356-62.

essarily presupposes these very institutions and values in the work of constructing the political conception of justice in the first place.

Let me put this point differently now. In *A Theory of Justice* there was a definite connection between Rawls's justificatory apparatus and such democratic institutions as majority rule: not only did ideas in the original position help specify the democratic institutions but these same ideas served to justify having such institutions. Thus, in *A Theory of Justice* the relevant public political culture was *defined* by reference to democratic political institutions (which, in turn, were specified by ideas in the original position construct). These institutions were the touchstone by which one identified a public political culture as democratic. But in *Political Liberalism* we are no longer able to do this. For there is now no device within Rawls's theory for independently or authoritatively specifying the democratic institutions. Instead, we make do with something that is much vaguer: the mere unstructured idea of a democratic political culture.

Now, my instinct here is that democracy crucially involves certain institutional features. If this is so, we do not really have a clear conception of democracy until these institutional features are well delineated and their positive relationship to other values and practices (themselves said to be democratic) established. Rawls has done neither of these things in *Political Liberalism* (neither the specifying nor the connecting). And because Rawls has done neither of these things, there is a significant gap in his theory.

My criticism so far is that Rawls's conception of democracy in *Political Liberalism* is not particularly well grounded, institutionally. Accordingly, it is not at all evident that his fundamental ideas (of society as a cooperative scheme for reciprocal benefit, of the citizen, of the well-ordered society, and so on) are themselves characteristically or peculiarly *democratic*.

Let me elaborate on this second point. One of Rawls's fundamental ideas is that citizens are free and equal. At first glance this might seem to be a paradigmatically democratic notion. However, the sense in which citizens are equal is underwritten by Rawls in a quite distinctive way; thus, what it means, in his account, for citizens to be equal largely depends on this particular grounding.

Rawls claims that there are two fundamental capacities or powers and, correspondingly, two "highest-order interests" of every individual citizen.<sup>31</sup> Thus, each person has, over that person's entire

31. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 47-88, 289-371.

life, (i) an interest in being able to have, formulate, revise, promulgate, live according to, and advance one's particular determinate conception of the good and (ii) an interest in exercising one's "sense of justice." That is, (to amplify this second point) each person has, over that person's entire life, an interest in living cooperatively with fellow citizens, on terms of mutual respect and reciprocal benefit, under a unified and stable scheme of basic political and economic institutions that has been organized by a shared set of principles of justice which each citizen can rationally affirm.

Citizens all have the two powers at a sufficient level to enable each of them to be fully contributing members of society over their entire lives. In having these powers at some such level, all the citizens are on the same footing. This, then, is the grounding idea behind Rawls's idea that the citizens are equal: they are equal in having reached what might be called this same *minimum* threshold level.<sup>32</sup>

Now, the idea that people have these two powers (or capacities) is by no means an idea peculiar to contemporary democratic thought. For that same idea would probably be found in the philosophy of Plato or Aristotle, in medieval social thought, as well as in democratic theory. Of course, there might be differences among these various theories as to whether all or almost all people could be "fully contributing" members of the body politic (with democrats inclined to emphasize that *all* are). The point here is, simply, that something like the two powers (and the attendant idea of equal *minimum* shares), though it might be found in democratic thought, would not be an idea that was especially and distinctively democratic.

The same is true for some of Rawls's other fundamental ideas. The idea of society as a cooperative venture for reciprocal benefit is not as such a distinctively democratic idea (except for special glosses

32. For the idea of a minimum threshold level, see *id.* at 19, 74, 79-80.

Rawls also uses the idea of the two powers or interests of the citizen to ground his elaboration of the concrete basic liberties that each citizen is to have equally. See Rawls, *Basic Liberties*, *supra* note 1, at 23-29, 47-51, 74; see also RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 310-15, 332-36, 358-59. Rawls offers liberties of conscience or freedom of personal association as examples of liberties justified under the *first* power or interest (the conception of the good interest) and freedom of political speech or of assembly as examples under the *second* power or interest (the sense of justice interest). Presumably, all of these liberties could be justified under *both* powers or interests. And some liberties (or protections from injury), for example, the due process rights to such things as a fair trial or the rights to bodily integrity, are justified as necessary to the full flourishing of the liberties justified in the "two fundamental cases." The basic liberties, authoritatively formulated, would be among the constitutional essentials. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 227. For further discussion of this derivation of the basic liberties, see MARTIN, *supra* note 23, at 45-61; James W. Nickel, *Rethinking Rawls's Theory of Liberty and Rights*, 69 *CHL-KENT L. REV.* 763 (1994).



that Rawls might put on the notion of reciprocity).<sup>33</sup> For this idea, like the idea of the powers or capacities of the citizen, is a mainstay of classical Greek and of medieval social thought. Nor, as before, is the idea of a well-ordered and stable society peculiarly democratic—for it is equally well embedded in ancient and medieval social thought (though, again, Rawls's emphasis on the importance of requiring publicity and public knowledge as to the principles and basic structural arrangements might lend itself to a democratic context more readily than to some others).

The point in all these cases is that Rawls's claim that the fundamental ideas of his new theory are latent in the public political culture of a contemporary democratic society seems thin, in that it lacks the needed democratic institutional focus. These may well be ideas that a contemporary democratic society would endorse but it is more likely they would endorse them as peculiarly *political* ideas than as democratic ones. For these ideas would probably not belong in a core characterization of what was distinctive about democratic political thought and practice (as against nondemocratic theories), or of what was distinctive of *contemporary* democracy (over against earlier democratic thought or practice).

One could remedy this signal defect by developing (or utilizing) a more robust theory of democracy than Rawls has used. Thus, (a) one could start with a characterization of the main contemporary democratic political institutions—for example, universal franchise (on a one person, one vote basis), contested voting, majority rule—and then, after exhibiting the coherence of this particular set of institutions, go on to show how certain attendant political values either reflect these institutions in some integral way (and are necessary to their correct working) or can be derived from them as characteristic and appropriate values for democratic citizens to hold.

Indeed, one could develop the idea of the original position—of the mediating model in *Political Liberalism*—out of this set of institutions. Here, features of the original position would be introduced only

33. Rawls distinguishes between "mutual advantage" and "reciprocity." RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 16-18 & n.18, 50, 54. The former takes as its benchmark "each person's present or expected future situation as things are [now]." *Id.* at 17. The latter (which Rawls prefers) takes as its fundamental point of comparison that "everyone benefits judged with respect to an appropriate benchmark of equality defined with respect to that world." *Id.* Unfortunately, Rawls leaves very vague exactly what this particular idea of equality amounts to. I have attempted to explicate this idea of the appropriate benchmark of equality, in the context of Rawls's difference or maximin principle, in section three of my paper *Economic Justice: Contractarianism and Rawls's Difference Principle*, in *SOCIAL CONTRACT: FROM HOBBS TO RAWLS* (David Boucher & Paul Kelly eds., forthcoming 1994).

insofar as they reflected salient details of the democratic institutions themselves. This would constitute a radically different procedure from the one Rawls used in *A Theory of Justice*, where the very reverse was the case. There the democratic institutions were themselves specified (and justified) as reflecting features of an antecedently described original position.

Or, alternatively to (a), one might (b) take the main contemporary democratic political institutions and then try to devise a plausible justification for those institutions by referring to *other* ideas or practices. Thus, for example, democratic procedures (as involving universal franchise, etc.) might be justified, at least in part, in the claim that they are a stable and reliable way of identifying, and then implementing, laws and policies that serve interests common to the voters or to a large number of them—presumably at least a majority. Of course, this claim, this rationale (if it is to count as a full and plausible justification) might need to be amplified in a variety of ways, and the amplifying material might include some of the attendant values referred to in the two previous paragraphs, in (a) above, and in the brief aside on the original position. But there would be nothing circular or question-begging about such a justification if the main rationale had already been established independently of any such use of attendant values in a justification. There might, then, be a place for Rawls's idea of society as a cooperative venture for reciprocal benefit, if it justificatorily supported, or helped support, the central rationale here.

The point here in both these cases is this: recognizably democratic institutions—such as universal franchise (on a one person, one vote basis), contested voting, majority rule—must ultimately serve as the touchstone of any account that claims to give a core characterization of what is distinctive about democratic political thought and practice (as over against nondemocratic theories). In both of the sample accounts I gave above, in (a) and (b), this was so. I think accordingly that Rawls's claim that the values he cites are peculiarly democratic ones would carry more conviction—and would appear less thin—if he were to undertake the sort of project just described, by taking the democratic political *institutions* as touchstone.

One way he could do this relatively comfortably in the theory he is actually advancing would be to identify the so-called constitutional essentials so as to include these democratic institutions.<sup>34</sup> Then, he

34. Thus Rawls argues that the principal step in developing a political conception of justice for a contemporary democratic society is to secure what he calls (following Kurt Baier) a "consti-

could use his fundamental ideas—on their own or in league with other ideas—to justify the having of those institutions. If we could, then, show that the main rationale (in the present case, Rawls's fundamental ideas) was initially independent of these institutions, then we would, in the fact of successful justification, have established a strong connection between those ideas and the democratic institutions. In so doing, we would have established the *democratic* credentials of the Rawlsian fundamental ideas. After that, they could be put to further use as grounds for constructing the principles of justice. And here, then, it would not appear arbitrary for Rawls to claim that he is using ideas latent in the political culture of a contemporary democratic society.

This brings me to the second main point I want to make in this section. One of the larger unresolved problems in Rawls's theory (one that runs from *A Theory of Justice* right through *Political Liberalism*) is actually establishing some of the main priorities that he so often cites. Thus, for example, Rawls needs to make a satisfactory case for putting basic constitutional rights (at least the ones he typically emphasizes: certain basic rights to liberty and to noninjury) *over* policies advancing corporate goods or other aggregate considerations. Or, more generally, for giving priority to the two principles over the common or public good, over "the net balance of advantages," and over efficiency.<sup>35</sup>

Since these priorities are part of the political conception of justice, they would need to be established in the first of Rawls's justifica-

tutional consensus." RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 158-68. But Rawls is very vague on institutional details here. My suggestion, then, would be that Rawls needs to include more such detail and to emphasize it in his elaboration of such a consensus.

One other point is worth noting here. Apparently, the "constitutional consensus" Rawls is contemplating specifically excludes certain elements in his second principle of justice; it excludes fair equality of opportunity (except for "freedom of movement and free choice of occupation") and the difference principle, though it does include some sort of commitment to a "social minimum." See *id.* at 228-9, 230, 232 n.14, 236 n.23. Thus, only a part of "basic justice"—only a part of the political conception of justice as initially conceived—is included within Rawls's proffered consensus on constitutional essentials. It would seem, then, that the price of such a consensus (in Rawls's view at least) is some watering down of the two principles of justice. If the constitutional consensus counts as the real core of Rawls's political conception (as it presumably does), then that same watering down becomes part of the price for obtaining "overlapping consensus" as well.

35. Rawls, *supra* note 6, at 562. See also RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 6, 156-57, 161, 175, 209, 223, 294; Rawls, *Priority of Right*, *supra* note 3, at 275-76; Rawls, *Overlapping Consensus*, *supra* note 3, at 17-18; Rawls, *supra* note 4, at 249-50; Rawls, *Basic Liberties*, *supra* note 1, at 8. Interestingly, in all these citations from *Political Liberalism*, Rawls has in mind, not the priority of the two principles *per se*, but only the priority of "certain basic rights, liberties, and opportunities" over "claims of the general good and of perfectionist values." RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 223; see also *id.* at 228-29, 232.

tory stages. And, again, merely to fetch them up by fiat out of the democratic political tradition, as latent or implicit there, seems both inherently doubtful and question-begging. One could, nonetheless, *argue* that these priorities indeed do exist there. The best way for Rawls to do so, as I have already suggested, is by utilizing the resources of a robust account of the main contemporary democratic political institutions. Thus, one might try to show that these institutions, when acting in accordance with what would justify them, require that *priority* be given to policies and laws in the interests of each and all—hence that priority be given to basic constitutional rights over other concerns of policy. Here the idea of priority would be built into Rawls theory side by side with the democratic institutions themselves.<sup>36</sup>

If some such robust account could be developed, with this result, then Rawls would be able to solve one of the longest standing problems in his theory (the problem of establishing certain relevant priorities). And he would have solved this problem in the way best for his new theory of justice, as part of developing an account of the main contemporary democratic institutions and of showing that certain values and norms did attend and were necessarily characteristic of those institutions or, alternatively, could be thought to justify them. And he would have done so at the right stage of his argument; for he would have established the relevant priorities as part of developing his political conception and, thus, would have accomplished the task of establishing those priorities *within* his political conception of justice, of constructing them before he went to the further justificatory stage, of overlapping consensus.

One final point seems worth making at this stage. Rawls may have overestimated the degree to which overlapping consensus is needed to provide stability to his political conception of justice (or overestimated the additional stability it could provide). Or he may have underestimated the divisive tendencies in the various comprehensive moral and religious doctrines and thus have been too taken with the idea that a consensus among them is even possible. We need to reckon, then, with the idea that in the end an overlapping consensus of such doctrines may not be possible nor, if possible, really neces-

36. I have tried to provide this more robust account by elaborating the contemporary notion of democracy and its institutions, and within that context the priority of basic constitutional rights. REX MARTIN, *A SYSTEM OF RIGHTS* 98-184, 303-22, 413-15 (1993). The main ideas here are conveniently summarized in my paper *Basic Rights*, *RECHTSTHEORIE*, 1993 Beiheft 15, at 191-201.

sary or particularly desirable. For these reasons it is important to make the political conception of justice *on its own* a focus of consensus and a basis of stability (by utilizing a robust account of the democratic institutions), without requiring the mediation of the various comprehensive doctrines.

### B. *Overlapping Consensus*

This said, we are now ready to probe that next stage of Rawls's theory a bit. Rawls's main claim, in talking of an overlapping consensus, is that the political conception, already established within the democratic political tradition, will also be endorsed and supported as the focus of a sort of consensus or agreement among the various comprehensive religious and moral doctrines. Rawls's language for expressing this support is sometimes quite misleading.

In his articles he wrote, for example, of the critical theories themselves as "premises" or "axioms" with respect to which the leading propositions of his political conception can be regarded as "theorems."<sup>37</sup> Such a characterization is not at all likely to be helpful. If critical moral theories are "true," itself a controversial and even doubtful claim, they are not true in the same way that, say, arithmetical axioms are true; that is, they are not *analytic* truths. In any event, the relationship between the critical theories, on the one hand, and the political conception, on the other, is *not* typically one of *logical* entailment.

But in *Political Liberalism*, Rawls by and large eschews loose talk of deductivity and settles instead for a far simpler claim.<sup>38</sup> Here his main claim is that *each* of the accredited theories of moral justification he refers to can afford its *own* reasons for accepting the two Rawlsian principles (and presumably the democratic constitutional essentials) as the principal feature of a political conception of justice—for accepting them as governing principles for the public domain of a given well-ordered society that has the sort of background and formation we find in contemporary democratic societies.<sup>39</sup>

37. See Rawls, *Overlapping Consensus*, *supra* note 3, at 9.

38. Rawls does, however, describe the relationship of Kantian comprehensive morality to the political conception as one of deduction. RAWLS, *POLITICAL LIBERALISM*, *supra* note 1, at 169, 171. And he does have one curious discussion of religion wherein God becomes a sort of axiom standing in a deductive relationship to the political conception. *Id.* at 242 & n.31.

39. Or, in another of his formulations, "citizens . . . view the political conception as derived from, or congruent with, or at least not in conflict with, their other values." *Id.* at 11, 140, 169. For further discussion see *id.* at 158-64, 168-72.

But even this notion of support may not do for one of the suggested theories of critical moral justification, Millian utilitarianism, that Rawls had in mind. For, I think it could be argued, though I will not do so here in any detail, that utilitarianism—including the theory of Mill—cannot support the priority of basic constitutional rights (at least the rights identified with the equal basic liberties) over corporate good or other aggregative welfare policies, if these policies can be adjudged to be preferable or even strongly supportable on utilitarian grounds.

Let me fill in behind this point a bit. Many people have thought that utilitarianism (including Mill's version) was somehow incompatible in principle with basic constitutional rights, or at least with their priority. The intuition that lies behind their view goes something like this.

They do not deny that the general happiness principle could support certain constitutional rights. For such rights are plausibly regarded as ways of acting or of being treated available to every person and justifiable by reference to the benefit of each and all. And, no doubt, some such policies for action (some such rights) would be justified as conducive to the general happiness. Rather, the problem these critics see is that no one can think that following any such policy or acting in accordance with any given right (especially if the *social rules* that formulate such things are kept fairly simple) will always and on every occasion yield up a result that is compatible with the general happiness principle. Sometimes deviating from the policy will have the greater welfare value. And, given the general happiness principle itself, the principle that the greater benefit should be preferred to the lesser and that normative requirements on action can always be set to achieve the greater benefit, that deviation *should* be taken.

It follows, then, that rights cannot have an overriding priority over corporate good or aggregate welfare considerations, insofar as these are supported (in the way just described) by the general happiness principle. Sometimes a right ought to yield to such considerations: it should do so where they hold the prospect of greater well-being. To deny this (once one accepts that sometimes deviations or exceptions or alternative policies can, all things considered, conduce to more net benefit than would setting up certain rights and keeping to them unfailingly) is to repudiate the foundational idea itself, the idea that a greater welfare is always to be preferred to a lesser and that action is always to be normatively regimented to that end.

The theory of "indirect" utilitarianism (found in Mill, advanced for a time by David Lyons and more recently by Wayne Sumner, and systematized by John Gray) is an attempt to deal with this problem. Roughly, these theorists assert that direct appeals to general welfare are self-defeating, all things considered, and that putting standing constraints on the principle—such as by installing a system of moral rules or a coherent set of civil rights justifiable by the standard of general happiness—is in fact productive of the greater well-being.<sup>40</sup>

Indirect utilitarians characteristically argue that, where rules conflict or rights do (as they inevitably will), some sort of appeal to the general happiness is in order. But building on what has already been said in favor of a constrained utilitarianism, they argue that the principle of general happiness should *not* directly determine what is to be done here; one does not say, for example, that action *A*, which comes under one rule, is to be done because on this occasion doing so produces the greater well being. Rather, the general happiness principle operates only indirectly in such a case. It bears down, not on individual actions *per se* (some of which might, as already noted, be described as deviating from the established rule) but on the rules themselves. It is used to determine which rule is weightier, all things considered and over the long run, or to help determine a policy (a rule of conduct), all things considered, for conduct when these particular moral rules (or these particular rights) conflict.

Thus, on their account it is possible to have policies for action (to have both moral rules and constitutional rights) that are justifiable by the standard of general happiness and, at the same time, to shield these policies from direct confrontation with (and possible overthrow by) the happiness principle on individual occasions. And it is possible to do so while still allowing these policies to remain sensitive to what produces the greater aggregate benefit on given, individual occasions—a sensitivity that is registered in the differential weights assigned the various rights and policies, an assignment that occurs gradually (over time and with experience) and cumulatively. Accordingly, indirect utilitarianism (if all its arguments and presumptions are allowed) seemingly establishes that utilitarianism is compatible with basic constitutional rights and their priority—at least in the case of

40. For a discussion, with extensive citation, of utilitarian attempts at a critical justification of rights and of utilitarian attempts to show that such rights would not, on utilitarian principles, be superseded by policies supported by considerations of general welfare or of the net balance of advantages, see MARTIN, *supra* note 36, at 324-29.

those rights that are themselves justifiable in accordance with the general happiness principle.

But I do not think that the indirect utilitarian strategy will work to save the *priority* of basic rights. Why? Well, we must begin by assuming that this strategy (like any good strategy) has a point, that one can offer a plausible reason for its being installed in the first place. Let us assume that the point here is provided by the idea of shielding moral rules and constitutional or basic rights from being overridden by corporate or aggregate political policies even when such policies were arguably supported as preferable by *direct* reference to the standard of general happiness on a given occasion. For it has to be granted that direct appeals could have this precise effect, if the indirect strategy is to have any point. But just because this is so, it follows that such appeals *could* gradually and cumulatively weight the scales against any given basic right and in favor of some nonrights policy. (Some policy that was neither a moral rule nor a basic right, nor straightforwardly derivable from such a rule or such a right.) Thus, the indirect strategy sometimes may be or can be unavailing in the project of protecting rights.

The matter here is empirical; it depends on facts of the world and not on any putative theoretical "truths" about utilitarianism. And when the strategy does prove unavailing, then a fundamental right must yield to the social policy that is supported, albeit indirectly, by considerations of *aggregate* well-being.

Philosophical utilitarianism, insofar as it remains true to what it has always been (that is, an aggregative theory) cannot plausibly allow a standing priority for distributive concerns (such as rights). It cannot allow a guarantee of benefits to individual persons in advance, so to speak, and across the board when that has the effect of tying the hands of utilitarian calculation against preferring corporate or aggregate policies when, cumulatively and all things considered, those policies could be seen to serve interests which conduce to greater benefit.

In this way, then, philosophical utilitarianism is incompatible with the notion of fundamental rights. It cannot plausibly provide a critical justification of a scheme of basic institutions in which constitutional rights have a settled priority over policies favoring corporate goods or aggregate welfare.<sup>41</sup>

41. Ted Vaggalis has suggested to me that utilitarianism, or at least Mill's utilitarianism (given its distinction of higher and lower pleasures and its preference for the higher), might escape this charge of incompatibility. But I cannot accept his claim. For Mill remains committed, even with the higher pleasures, to the basic utilitarian idea of preferring the greater amount



Since the factor of priority is very important to Rawls's political conception, this defect (if I may call it that) of Millian utilitarianism is decisive. It seems, then (even if we assume the favored case of indirect utilitarianism), that utilitarians cannot accept, let alone justify, Rawls's account of constitutional rights. And it seems too that Rawls's second pattern of justification might well be seriously disrupted, or even overthrown, by a utilitarian secession from the overlapping consensus (over the key issue of the priority of basic rights).

Of course, Millian indirect utilitarianism could probably support a political conception *like* Rawls's own (could, for example, support one that was like Rawls's first principle in every respect except that the supported conception lacked the crucial feature of priority—of the basic liberty rights over policies supported by considerations of general happiness). But such support is beside the point. For the fact remains that Millian utilitarianism cannot support the specific political conception of justice that Rawls advocates and with which his name is justly and indelibly identified.

It could, of course, still be true that Kantian moral philosophy might provide support, the support of an accredited critical moral theory, for Rawls's political conception. But I cannot confidently say that it would, for largely Rawls merely asserts that (like Millian utilitarianism) it would; but he nowhere exhibits this support in anything like a rigorous fashion. In any event, even if such support was there (and could convincingly be shown to be so), more than mere Kantian support would be required if we were to have an "overlapping consensus" and were to avoid thereby the partisanship and controversy of endless speculative disquisitions between and among the reasonable moral and religious theories.

We reach, in the end, a rather unsatisfactory result for Rawls's theory of justice. In the form his theory had initially, in 1971, it was not an acceptable or creditable theory of critical moral justification. And in the form it has had more recently, after 1980 and especially since 1985, it does not represent itself as a critical moral theory at all. It is not clear, moreover, that Rawls's theory would be supported by

of pleasure to the lesser. See Rex Martin, *A Defence of Mill's Qualitative Hedonism*, 47 *PHILOSOPHY* 140, 146, 148. In any event the claim Vaggalis makes is beside the point: for Rawls in *Political Liberalism* is not restricting the kind of utilitarianism that can join the overlapping consensus to the Millian variety; he explicitly says there that classical utilitarianism can join in too (and will). But the utilitarianism of Bentham and Sidgwick makes no distinction of higher and lower pleasures and, hence, could not be saved by that distinction from holding views incompatible with Rawls's idea of the priority of basic rights (which in turn is at the heart of the political conception of justice on which overlapping consensus is to focus).

any of the accredited theories of critical moral justification that he names, and, contrary to his express expectation, very unlikely that it would be in one such case—that of utilitarianism, even of the Millian stamp.

